DOCUMENT RESUME

ED 078 052 TM 002 867

AUTHOR Kristof, Walter
TITLE On Accuracy in Reliability Estimation.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RB-73-24
PUB DATE Apr 73
NOTE 12p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Expectation; Sampling; Scores; *Statistical Analysis; Tables (Data); Technical Reports; *Test Reliability; *Theories

ABSTRACT

This study in parametric test theory deals with the statistics of reliability estimation when scores on two parts of a test follow a binormal distribution with equal (case 1) or unequal (case 2) expectations. In each case biased maximum-likelihood estimators of reliability are obtained and converted into unbiased estimators. Sampling distributions are derived. Second moments are obtained and utilized in calculating mean square errors of estimation as a measure of accuracy. A rank order of four estimators is established. There is a uniformly best estimator. Tables of absolute and relative accuracies are provided for various reliability parameters and sample sizes. (Author)

RESEARCH BULLETIN

RB-73-24

ON ACCURACY IN RELIABILITY ESTIMATION

Walter Kristof

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the author. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
April 1973

ON ACCURACY IN RELIABILITY ESTIMATION

Abstract

This study in parametric test theory deals with the statistics of reliability estimation when scores on two parts of a test follow a binormal distribution with equal (case 1) or unequal (case 2) expectations. In each case biased maximum-likelihood estimators of reliability are obtained and converted into unbiased estimators. Sampling distributions are derived. Second moments are obtained and utilized in calculating mean square errors of estimation as a measure of accuracy. A rank order of four estimators is established. There is a uniformly best estimator. Tables of absolute and relative accuracies are provided for various reliability parameters and sample sizes.

ON ACCURACY IN RELIABILITY ESTIMATION

1. Introduction and Preliminaries

The present paper constitutes a part of a larger study in parametric test theory. It is devoted to the development of formulas relevant in reliability estimation with emphasis on the accuracy of various estimates when explicit distributional assumptions regarding test scores are made. This approach also permits the derivation of statistical tests of hypotheses about reliability parameters. It does not limit itself to giving conjectured large sample estimators and stating relationships between parameters.

Probably the first paper with emphasis on small sample distributions of estimated reliabilities was written by Kristof [1963]. Feldt [1965] also took up the subject. Other papers in the area of parametric test theory were given by Kristof [1970, 1972].

Inferences about the reliability of a given test require repeated measurement in one form or another on a sample of subjects. Two approaches to data collection are common: (a) one obtains multiple measurements using basically the same test whose reliability is the quantity of interest; (b) one obtains multiple measurements using comparable parts of the test whose reliability is the quantity of interest.

In the second case the reliability of the component parts is stepped up to give the reliability of the total test. This procedure is not required in the first case. One might, therefore, assume that case (b) should lead to a statistical theory more complicated than that based on case (a). However, the opposite is true. Not much work with emphasis on statistics has been presented for case (a). There is an important paper by Olkin and Pratt [1958] that bears on some of the existing problems. Contributions dealing with case (b) are more numerous.

It is case (b) to which we address ourselves here also. Some of the results will be special cases of formulas in Kristof [1963]; others will be new. In particular, we wish to include a table of mean square errors of reliability estimators so that the comparative merits of several estimators can be assessed. It is obvious that the choice of a particular estimator out of a number of possible ones could well be based upon a comparison of accuracies.

In practice the most important instance of case (b) occurs when a test has been divided into just two parts. Using the classical test theory model and denoting total observed, true and error scores by X, T and E, respectively, and corresponding scores on the parts by X_i, T_i and E_i, $i = 1,2$, we write

$$(1) \qquad X = X_1 + X_2 = T + E = T_1 + T_2 + E_1 + E_2 \quad .$$

Let us perform the transformation

$$(2) \qquad Y_1 = X_1 - X_2 \quad , \quad Y_2 = X_1 + X_2 \quad .$$

As to second moments classical test theory tells us that

$$(3) \qquad \sigma^2_{Y_1} = \sigma^2_E \quad , \quad \sigma^2_{Y_2} = \sigma^2_T + \sigma^2_E \quad , \quad \sigma_{Y_1Y_2} = 0 \quad .$$

It has been assumed that the two parts are indistinguishable as regards true score variance and mean and variance of errors of measurement. It has not been assumed that true scores on the parts have equal means. If they do, we will speak of case 1; if they do not, we have case 2.

In each case the reliability of the total test is given by

$$(4) \qquad \rho = \sigma_T^2/(\sigma_T^2 + \sigma_E^2) = 1 - \sigma_{Y_1}^2/\sigma_{Y_2}^2 \quad .$$

If we supply a hat to a parameter to denote its maximum-likelihood estimator we obtain

$$(5) \qquad \hat{\rho} = 1 - \hat{\sigma}_{Y_1}^2/\hat{\sigma}_{Y_2}^2 \quad .$$

At this point it is necessary to introduce specific distributional assumptions. It will be assumed that X_1, X_2 follow a bivariate normal distribution.

2. Distribution of $\hat{\rho}$

Case 1. Since Y_1 has expectation zero, we get $\hat{\sigma}_{Y_1}^2 = \sum_i y_{1i}^2/N$ where y_{1i} signifies the observed difference score for subject i, N being the sample size. Further, $\hat{\sigma}_{Y_2}^2 = \sum_i (y_{2i} - \bar{y}_2)^2/N$ where y_{2i} is the observed value of Y_2 for subject i and $\bar{y}_2$ is the arithmetic mean over subjects. Then $\hat{\rho}$ is given by (5). Quantity $\hat{\sigma}_{Y_2}^2$ is biased, $\hat{\sigma}_{Y_1}^2$ is not. Replacing N by $N - 1$ in $\hat{\sigma}_{Y_2}^2$ yields the usual unbiased variance estimator. Combination of (4) and (5) gives

$$(6) \qquad F_{N,N-1} = (N - 1)(1 - \hat{\rho})/N(1 - \rho) \quad .$$

This magnitude is distributed as F with $df_1 = N$, $df_2 = N - 1$. It would be a simple matter to obtain the distribution of $\hat{\rho}$ explicitly from (6). However, (6) is sufficient for our purposes since it allows us to test hypotheses about ρ and to establish confidence intervals for ρ.

Case 2. Now we have $\hat{\sigma}^2_{Y_1} = \sum_i (y_{1i} - \bar{y}_1)^2/N$, $\hat{\sigma}^2_{Y_2}$ as before. Both $\hat{\sigma}^2_{Y_1}$ and $\hat{\sigma}^2_{Y_2}$ are biased. This is immaterial, however, because the same numerator appears in both formulas. We find that the quantity

$$(7) \qquad F_{N-1,N-1} = (1 - \hat{\rho})/(1 - \rho)$$

follows an F distribution with $df_1 = df_2 = N - 1$, $\hat{\rho}$ defined in (5). It is possible to replace (7) by an equivalent but possibly more convenient formula if we use the fact that

$$(8) \qquad t_n = \sqrt{n}\ (\sqrt{F_{n,n}} - 1/\sqrt{F_{n,n}})/2$$

follows a t distribution with $df = n$ when $F_{n,n}$ follows an F distribution with $df_1 = df_2 = n$. This relationship was discovered independently by Aroian [1953], Cacoullos [1965] and Kristof [1972]. Thus

$$(9) \qquad t_{N-1} = (\rho - \hat{\rho})\sqrt{N-1}\,/2\sqrt{(1-\rho)(1-\hat{\rho})} \quad .$$

Equation (9) has no equally simple counterpart in case 1.

3. Bias of $\hat{\rho}$

Case 1. Taking expectations on both sides of (6) we get [Kendall & Stuart, 1958, p. 378]

(15) $$F_{N-1,N-1} = (N - 1)(1 - \tilde{\rho})/(N - 3)(1 - \rho) \quad .$$

Again $\tilde{\rho}$ is minimum variance unbiased.

5. Variance of $\hat{\rho}$

Case 1. Knowledge of the variance of $\hat{\rho}$, $\sigma^2_{\hat{\rho}}$, will enable us to calculate the mean square error of estimation in $\hat{\rho}$. Taking variances on both sides of (6) we get [Kendall & Stuart, 1958, p. 378]

(16) $$\sigma^2_{\hat{\rho}} = 2N(2N - 3)(1 - \rho)^2/(N - 3)^2(N - 5) \quad .$$

Case 2. From (7) we obtain

(17) $$\sigma^2_{\hat{\rho}} = 4(N - 1)(N - 2)(1 - \rho)^2/(N - 3)^2(N - 5)$$

by an analogous operation.

6. Mean Square Errors of Estimation in $\hat{\rho}$ and $\tilde{\rho}$

The mean square error of estimation, MSE, is a likely choice if a measure of accuracy of an estimator is sought. MSEs of $\hat{\rho}$ are found when we use

(18) $$MSE = \mathcal{E}(\hat{\rho} - \rho)^2 = \sigma^2_{\hat{\rho}} + (\mathcal{E}\hat{\rho} - \rho)^2$$

with $\hat{\rho}$ replaced by $\tilde{\rho}$ if the latter quantity is of interest. In this case MSE = $\sigma^2_{\tilde{\rho}}$.

Magnitudes $\sigma^2_{\hat{\rho}}$ in cases 1 and 2 are given in (16) and (17). $(\mathcal{E}\hat{\rho} - \rho)^2$ is obtained from (10) and (11). Quantities $\sigma^2_{\tilde{\rho}}$ may be determined from (12) and (14).

After calculating the four mean square errors of estimation we wish to summarize the results. In each case the MSE is the product of $(1 - \rho)^2$ and a factor which depends only on N. These factors are listed in the following fourfold table:

	Case 1	Case 2
MSE of $\hat{\rho}$	$(4N + 15)/(N - 3)(N - 5)$	$4(N + 1)/(N - 3)(N - 5)$
MSE of $\tilde{\rho}$	$2(2N - 3)/N(N - 5)$	$4(N - 2)/(N - 1)(N - 5)$

For conciseness let us use the symbol MSE_{ij} to indicate the mean square error of estimation in case i, $i = 1,2$, for estimator of type j where $j = 1$ refers to $\hat{\rho}$ and $j = 2$ to $\tilde{\rho}$. We see that

$$(19) \qquad MSE_{12} < MSE_{22} < MSE_{21} < MSE_{11}$$

for all $\rho < 1$ and $N > 5$. Hence, if accuracy of estimation is the criterion, $\tilde{\rho}$ in case 1 is best and $\hat{\rho}$ in case 1 is worst. In addition, $\tilde{\rho}$ has the advantage of being unbiased. If the division of the total test into two parts is such that only case 2 applies, then again $\tilde{\rho}$ is preferred to $\hat{\rho}$. At any rate, $\hat{\rho}$ in case 1 is the poorest possible choice because MSE_{11} is largest.

In the following table we list the ratios MSE_{ij}/MSE_{12}, $ij \neq 12$, for selected sample sizes rounded to four decimal places:

	$N = 6$	10	20	50	100
MSE_{22}/MSE_{12}	1.0667	1.0458	1.0242	1.0099	1.0050
MSE_{21}/MSE_{12}	3.1111	1.8486	1.3355	1.1187	1.0571
MSE_{11}/MSE_{12}	4.3333	2.3109	1.5103	1.1790	1.0859

For small sample sizes $\hat{\rho}$ is generally quite inferior. If sample size exceeds 100, the differences tend to be negligible, however. This table can be easily extended if we use the relations

$$MSE_{22}/MSE_{12} = 1 + (N - 3)/(2N^2 - 5N + 3)$$

$$MSE_{21}/MSE_{12} = 1 + (11N - 9)/(2N^2 - 9N + 9) \tag{20}$$

$$MSE_{11}/MSE_{12} = 1 + (33N - 18)/(4N^2 - 18N + 18) \quad .$$

For the best estimator, $\tilde{\rho}$ in case 1, we give the actual values of MSE_{12} for various ρ and N rounded to four decimal places:

	$N = 6$	10	20	50	100
$\rho = .60$	.4800	.1088	.0395	.0138	.0066
.70	.2700	.0612	.0222	.0078	.0037
.80	.1200	.0272	.0099	.0034	.0017
.90	.0300	.0069	.0025	.0009	.0004
.95	.0075	.0017	.0006	.0002	.0001

This table reflects the rapid gain of accuracy of estimation as N and/or ρ increase. We see in particular that the accuracy of estimation will be uniformly high if $N \geq 50$ and $\rho \geq .60$. It will be interesting to observe the trade-off occurring between ρ and N. For instance, the accuracies in estimating $\rho = 0.60$ when $N = 100$ and $\rho = 0.90$ when $N = 10$ are about equal.

REFERENCES

Aroian, L. A. A certain type of integral. American Mathematical Monthly, 1953, 50, 382-383.

Cacoullos, T. A relation between t and F -distributions. Journal of the American Statistical Association, 1965, 60, 528-531.

Feldt, L. S. The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty. Psychometrika, 1965, 30, 357-370.

Kendall, M. G. & Stuart, A. The advanced theory of statistics. Vols. 1, 2. New York: Hafner, 1958, 1961.

Kristof, W. The statistical theory of stepped-up reliability coefficients when a test has been divided into several equivalent parts. Psychometrika, 1963, 28, 221-238.

Kristof, W. On the sampling theory of reliability estimation. Journal of Mathematical Psychology, 1970, 7, 371-377.

Kristof, W. On a statistic arising in testing correlation. Psychometrika, 1972, 37, 377-384.

Olkin, I. & Pratt, J. W. Unbiased estimation of certain correlation coefficients. The Annals of Mathematical Statistics, 1958, 29, 201-211.